

Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data

Blossom H. PATTERSON, C. Mitchell DAYTON, and Barry I. GRAUBARD

High fruit and vegetable intake is associated with decreased cancer risk. However, dietary recall data from national surveys suggest that, on any given day, intake falls below the recommended minima of three daily servings of vegetables and two daily servings of fruit. There is no single widely accepted measure of "usual" intake. One approach is to regard the distribution of intake as a mixture of "regular" (relatively frequent) and "nonregular" (relatively infrequent) consumers, using an indicator of whether an individual consumed the food of interest on the recall day. We use a new approach to summarizing dietary data, latent class analysis (LCA), to estimate "usual" intake of vegetables. The data consist of four 24-hour dietary recalls from the 1985 Continuing Survey of Intakes by Individuals collected from 1,028 women. Traditional LCA based on simple random sampling was extended to complex survey data by introducing sample weights into the latent class estimation algorithm and by accounting for the complex sample design through the use of jackknife standard errors. A two-class model showed that 18% do not regularly consume vegetables, compared to an unweighted estimate of 33%. Simulations showed that ignoring sample weights resulted in biased parameter estimates and that jackknife variances were slightly conservative but provided satisfactory confidence interval coverage. Using a survey-wide estimate of the design effect for variance estimation is not accurate for LCA. The methods proposed in this article are readily implemented for the analysis of complex sample survey data.

KEY WORDS: Categorical data; Cluster sample; Design effect; Dietary propensity scores; Jackknife; Latent class model; Sample weight.

1. INTRODUCTION

Frequent consumption of fruit and vegetables has been linked to reduced cancer incidence. For many cancer sites, persons with low intake of these foods experience about twice the cancer risk as do those with high intake (Block, Patterson, and Subar 1992). Dietary surveillance is used to monitor the intake of foods that are important risk factors for cancer and heart disease. A major goal of dietary surveillance is to estimate the distribution of intake of nutrients and foods in the population. At a policy level, information on dietary intake is important for shaping dietary guidance and for the evaluation of dietary intervention programs, such as the national Five A Day program, that encourages the consumption of five or more servings of fruits and vegetables daily (Subar et al. 1994). In this article we focus on vegetable consumption alone.

Twenty-four-hour dietary recall data from national surveys suggest that, on any given day, consumption falls below the recommended three or more daily servings of vegetables (Patterson, Block, Rosenberger, Pee, and Kahle 1990; Patterson, Harlan, Block, and Kahle 1995; Krebs-Smith, Cook, Subar, Cleveland, and Friday 1995). The mean of two non-consecutive recall days from the 1994–1996 Continuing Survey of Food Intakes by Individuals (CSFII) showed that 55% of the population age 20 years and older consumed three or more servings of vegetables (U. S. Department of Agriculture 1998). A goal of *Tracking Healthy People 2010* (U. S. Department of Health and Human Services 2000) is increasing consumption to 75%. New dietary assessment methods

that include estimation of the regularity of consumption are critical to measure progress toward this goal.

Although there is no clear definition of "usual" dietary intake (Guenther 1997), it can be regarded as intake over some long period. Methods currently used for measuring usual intake have been described by Thompson and Byers (1994). One method, the focus of this study, requires the collection of two or more 24-hour recalls or daily food diaries. Several methods for combining dietary records have appeared in the literature. The 1977–1978 Nationwide Food Consumption Survey (Human Nutrition Information Service 1983) estimated the percentage of individuals using a particular food as the number reporting consuming that food at least once in the 3-day survey period, divided by the group size. Hartman et al. (1990) reported mean daily intake for several food groups based on 12 two-day diaries. Popkin, Siega-Riz, and Haines (1996) summarized information on consumption of various foods from a single 24-hour recall into a dietary score for each respondent.

Analyses of dietary data have focused primarily on nutrient intake (e.g., fat, vitamin A) rather than on the intake of particular foods (e.g., butter, carrots). Nutrients are typically consumed daily in some quantity, with the result that zero intakes rarely occur. In contrast, specific foods are consumed less frequently, and zero counts are expected to occur. In the 1985 CSFII dataset, the intake of each food consumed by a respondent was reported in grams based on portion-size estimates (U. S. Department of Agriculture 1987). Thus, for a given food, either an amount in grams or a 0 is associated with each respondent for each recall day. The distribution of intake for a given food consists of a continuous component that can be modeled by, for example, a lognormal distribution with a point mass at 0. Alternatively, the nonzero values can be classified into a set of ordered categories. The data also can be treated

Blossom H. Patterson is Mathematical Statistician, Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD 20892 (E-mail: bp27z@nih.gov). C. Mitchell Dayton is Professor, Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park, MD 20742. Barry I. Graubard is Senior Investigator, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892. This work was part of the first author's Ph.D. dissertation in the Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park. The authors thank Susan Krebs-Smith for her insightful comments on nutritional issues, Kevin Dodd for his interpretation of our model in terms of propensity scores, and the referees for their suggestions that greatly improved the manuscript.

© 2002 American Statistical Association
Journal of the American Statistical Association
September 2002, Vol. 97, No. 459, Applications and Case Studies
DOI 10.1198/016214502388618465

as counts or “mentions” of a food or foods in a food group, where a mention is any nonzero quantity. The distribution of intake can then be modeled by, for example, a Poisson distribution with overdispersion at 0 (Smith, Graubard, and Midthune 1997) or by other mixture models. In all of these cases, the distribution can be regarded as arising from a mixture of “regular,” that is, relatively frequent consumers and “nonregular,” that is, relatively infrequent consumers.

Dietary data also have been analyzed separating consumers and nonconsumers (Subar et al. 1993; Patterson et al. 1995). Such data can be further simplified by identifying a consumer of a given food (or food in a food group) with a 1 and a non-consumer with a 0. These two approaches have the advantage of mitigating the measurement error inherent in dietary data based on portion size (Smith 1991; Young and Nestle 1995). Reporting the fact of consumption is simpler and likely to be more accurate than the quantity consumed. In fact, such measurement error may be a reason for dichotomizing the data.

Latent class analysis (LCA) is a method of grouping individuals with respect to some underlying, unobservable variable based on data from polytomous indicators or items. This method can be useful in the analysis of the intake of foods, for example, in estimating the regularity of vegetable consumption. Individuals in a sample can be classified into two or more latent classes based on binary data reflecting their consumption/nonconsumption of vegetables.

National dietary data are typically collected in surveys that have complex sample designs involving multistage sampling with sample weighting. The analysis of such designs has been described for mixture models (Wedel, ter Hofstede, and Steenkamp 1998; Patterson 1998). The focus of this study was to fit a population latent class model (LCM) to data from the 1985 CSFII and to develop appropriate LCA methods for complex sample surveys (see Patterson 1998). In Section 2 the CSFII is summarized. In Section 3 the LCM is introduced, and the jackknife is presented as a method of estimating standard errors for the LCM parameters. The CSFII data are analyzed in Section 4, and a simulation is presented in Section 5. Finally, the method and results are discussed in Section 6.

2. THE CONTINUING SURVEY OF FOOD INTAKES BY INDIVIDUALS

The 1985 CSFII comprised a multistage stratified area probability sample of women age 19–50 living in private households in the 48 conterminous states. The conterminous United States was divided into 60 “relatively homogeneous” strata, and 2 primary sampling units (PSUs) were sampled per stratum. Although the survey was designed to be self-weighting, differential sample weights were computed to reflect various levels of nonresponse at the household and individual levels. (For more details see U. S. Department of Agriculture 1987.)

In an attempt to estimate usual intake, six dietary recalls of foods consumed during the previous 24 hours were collected at about 2-month intervals. The first recall was collected in a face-to-face interview; the next five recalls were done by telephone. The public-use CSFII data tape includes all women who participated in the face-to-face interview and completed at least three phone recall interviews. For women who completed the face-to-face interview and four or five phone recall

Table 1. Distribution of Days on Which Respondents Reported Eating a Vegetable

| Number of Days | Weighted Percent | Cumulative Weighted Percent |
|----------------|------------------|-----------------------------|
| 0 | 3.1 | 3.1 |
| 1 | 9.4 | 12.5 |
| 2 | 22.4 | 34.9 |
| 3 | 37.7 | 72.6 |
| 4 | 27.4 | 100.0 |

interviews, three phone recalls were randomly selected. Thus recalls 2–4 do not represent the same recall occasions for all of the women. Those women who were lost because of insufficient numbers of interviews were accounted for in the sample weights.

This dataset has been used as an exemplar to test various methods of analysis because it consists of four independent food records on each respondent (Haines, Hungerford, Popkin, and Guilkey 1992; Nusser, Carriquiry, Dodd, and Fuller 1996). The dataset used in this analysis consists of 1,028 women who had nonzero food intake on all of the recall days. (Four women who had zero food intake on at least one of the recall days were eliminated.) Five of the strata in the dataset had a single PSU. For the purposes of variance estimation, these were paired/combined in such a way that the resulting 56 strata each contained 2 PSUs and 1 stratum contained 3 PSUs.

For each interview, a respondent was assigned a value of 1 if she reported consuming any vegetable on the recall day (i.e., one or more mentions) and a 0 otherwise. This broad group of vegetables includes salad, legumes, and such foods as peas, carrots, corn, and other green and deep-yellow vegetables, but not potatoes; this group of vegetables is of special interest because of its nutrient content.

The weighted relative distribution for the number of recall days on which sampled women reported consuming at least one vegetable is shown in Table 1. On average, respondents reported consuming at least one vegetable on 2.8 days out of 4. Approximately 73% of respondents did not consume a vegetable on at least 1 of the 4 recall days and 12.5% did so on at most 1 of the 4 days.

3. LATENT CLASS MODEL FOR SURVEY DATA

An LCM is used to explain underlying, unobservable categorical relationships, or latent structures, that characterize discrete multivariate data (Lazarsfeld and Henry 1968; Goodman 1974; Dayton and Macready 1976; Haberman 1979). When food intake is dichotomized, LCA is a technique uniquely suited to combining dietary information from several food records or 24-hour recalls to characterize the regularity of vegetable consumption of a population (as here) or population subgroup for a food or food group of interest. Here, regularity of vegetable consumption is the underlying structure of interest.

Methods for LCA that take into account sample design features, such as sample weighting, clustering, and stratification used in complex surveys like the CSFII, have not been described in the literature. However, results from regression

analysis have shown that if data are collected under a complex sampling design and simple random sampling (SRS) is assumed in the analysis, then parameter estimates can be biased and standard errors underestimated (Korn and Graubard 1999, pp. 159–172).

Let $\mathbf{Y}_i = \{y_{ij}\}$ be the vector-valued response for J survey items, $j = 1, \dots, J$, for the i th respondent drawn from a finite population of size N . The polytomous response options take on discrete values $r = 1, \dots, R_j$ for the j th item. The probabilities, θ_l , for the unobserved LCs, c_l , $l = 1, \dots, L$, are called LC proportions. Item-conditional probabilities, $\alpha_{l|j1} \dots \alpha_{l|jR_j}$, represent the probabilities of response r to item j given membership in LC l . Thus for each item j there is an R_j -vector of conditional probabilities. To illustrate the notation, consider data based on four polytomous variables with, say, $R_1 = 2$, $R_2 = 4$, $R_3 = 3$, and $R_4 = 2$ denoting the number of discrete values taken on by each item. $\mathbf{Y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})' = (1, 3, 2, 2)'$ might represent the responses for the i th respondent.

The traditional LCM can be defined as

$$\Pr(\mathbf{Y}_i | c_l) = \prod_{j=1}^J \prod_{r=1}^{R_j} \alpha_{l|jr}^{\delta_{ijr}} \quad (1)$$

and

$$\Pr(\mathbf{Y}_i) = \sum_{l=1}^L \theta_l \Pr(\mathbf{Y}_i | c_l), \quad (2)$$

where the Kronecker delta is defined as

$$\delta_{ijr} = \begin{cases} 1 & \text{iff } y_{ij} = r, \quad r = 1, \dots, R_j \\ 0 & \text{otherwise.} \end{cases}$$

The usual restrictions for item-conditional probabilities apply (i.e., $\sum_{r=1}^{R_j} \alpha_{l|jr} = 1 \forall j$) and, in addition, the LC proportions sum to 1 (i.e., $\sum_{l=1}^L \theta_l = 1$). Note that, unlike models proposed by Clogg and Goodman (1984, 1985), the model in (1) and (2) does not directly reflect grouping of respondents, for example, males and females.

In the context of the CSFII dataset, the response variables, y_{ij} , are (dichotomous) indicators of consumption of some food of interest on each of 4 recall days. For a two-class model, the LC proportions refer to the proportions in "regular" and "non-regular" vegetable consumption groups. Each item-conditional probability refers to the probability or "propensity" for consuming at least one vegetable on the corresponding recall day, given membership in a specific consumption group.

Assuming SRS with a sample of size n , the log-likelihood is

$$\Lambda = \sum_{i=1}^n \ln \sum_{l=1}^L \theta_l \Pr(\mathbf{Y}_i | c_l) = \sum_{i=1}^n \ln \left\{ \sum_{l=1}^L \theta_l \prod_{j=1}^J \prod_{r=1}^{R_j} \alpha_{l|jr}^{\delta_{ijr}} \right\}. \quad (3)$$

Fundamental to classical LCA is the assumption that the observed variables are independent within LCs. Parameter estimation can be accomplished by means of maximum likelihood methods using conventional iterative algorithms such as Newton–Raphson or the EM algorithm (Dempster, Laird, and Rubin 1977; Heinen 1996).

In the case of non-SRS, where sample weights, w_i , are available for each respondent (e.g., from a public-use data tape), these weights are usually the product of the reciprocal of the sample inclusion probability, a factor that adjusts for nonresponse, and a factor that reflects poststratification adjustment. These weights may be expansion weights that sum to the total population size or relative weights that are scaled to sum to the sample size. A sample-weighted pseudo-log-likelihood can be defined as

$$\begin{aligned} \Lambda_w &= \sum_{i=1}^n w_i \ln \sum_{l=1}^L \theta_l \Pr(\mathbf{Y}_i | c_l) \\ &= \sum_{i=1}^n w_i \ln \left\{ \sum_{l=1}^L \theta_l \prod_{j=1}^J \prod_{r=1}^{R_j} \alpha_{l|jr}^{\delta_{ijr}} \right\}. \end{aligned} \quad (4)$$

Maximizing the pseudo-log-likelihood simultaneously with respect to θ_l and $\alpha_{l|jr}$ provides design-consistent estimates of the underlying population parameters (Pfeffermann 1993).

The LC model used in this article can be expressed as a log-linear model, using either a Poisson or binomial distribution for the cell counts in the finite population. In the survey setting, the weighted pseudo-likelihood is obtained by replacing the unweighted cell counts with the sample-weighted cell counts in the likelihood, as implied by (4). An alternative method for log-linear analysis of sample weighted contingency tables (Clogg and Eliason 1987; Agresti 1990, p. 199) uses the unweighted cell counts with an offset, consisting of the log of the inverse of the average cell sample weight, in each cell of the contingency table. Under a correctly specified log-linear model for the population, this method will produce consistent estimates of model parameters, that is, estimates that are asymptotically equal to the values that would have been obtained had they been computed using the entire finite population. However, if the log-linear model for the population is misspecified, then the two methods will not agree asymptotically. We prefer the weighted pseudolikelihood method because its estimates will be approximately unbiased for values of the population model parameters, regardless of whether the model was correctly specified.

Although not explicit in the models as written, clustering is taken into account when estimating standard errors. Cluster sampling, such as that in the CSFII, induces correlation among responses and typically results in sampling variances that are larger than would be the case under SRS. Further, standard test statistics (such as the Pearson chi-squared) used in LCA are no longer asymptotically distributed as chi-squared random variables when the data arise from a survey with clustered sampling (Hidiroglou and Rao 1987; Roberts, Rao, and Kumar 1987).

Two methods of calculating standard errors for complex sample survey data are adjustment using a design effect (deff), and the use of a replication method such as the jackknife. The first method was used in LCA by Haertel (1984a, 1984b, 1989). The jackknife is applicable to virtually any type of complex sample design (Wolter 1985) and is known to provide reasonable standard errors for many statistics that are smooth (differentiable) functions of the data (Efron 1982). The applicability of the jackknife to the estimation of LC parameters under SRS has been studied empirically by Bolesta (1998).

In complex sample surveys, sample weights and clustering usually inflate the variance, whereas stratification may result in variance reduction. The deff, the ratio of the variance under the full design to the variance assuming SRS, is usually greater than 1. Kish (1965, pp. 258–259) described this “comprehensive factor” as attempting to summarize the effects of “various complexities in the sample design, especially those of clustering and stratification. . . (and) may include effects of . . . varied sampling fractions.” He noted that the design effect can be used to obtain an effective sample size, $n' = \frac{n}{deff}$, to be used in place of n , the actual sample size, in the calculation of standard errors. The size of the deff depends on the variable being estimated and may vary among subsets of the population. If a surveywide estimate of the deff is available and applied to all estimates, then the adjustment may be too large or too small and also may give misleading results for population subgroups (Korn and Graubard 1995a). Haertel (1984a, 1984b, 1989) took the sample design into account in the calculation of standard errors in LCA by using an external estimate of the overall deff to estimate an effective sample size, which he then used in the calculation of standard errors.

The jackknife was introduced as a method of bias reduction by Quenouille (1949), and the procedure was subsequently used to estimate the variance of a parameter estimate (Mosteller and Tukey 1968; Miller 1968, 1974). Frankel (1971) made an early application of this technique to complex sample survey data. The method proceeds as follows. Let $\hat{\gamma}$ be the sample-weighted estimate of a population parameter of interest, γ , for a sample of size n . In a complex sample survey with stratification and clustering, the PSUs are randomly grouped within strata, where each random group has approximately the same number of PSUs. Let k_h denote the number of random groups in stratum h , $h = 1, \dots, H$. A random group of PSUs in stratum h is omitted, and the remaining observations in that stratum are reweighted by a multiplicative factor $k_h/(k_h - 1)$. The usual parameter estimates, called *jackknife estimates*, are derived from the reduced sample. This process is repeated sequentially for the entire sample of PSUs. A variance estimator based on the jackknife is (Wolter 1985)

$$\widehat{\text{var}}^J(\hat{\gamma}) = \sum_{h=1}^H \left[\sum_{s=1}^{k_h} \frac{k_h - 1}{k_h} (\hat{\gamma}_{(sh)} - \hat{\gamma})^2 \right], \tag{5}$$

where $\hat{\gamma}_{(sh)}$ is the jackknife estimate of γ omitting group s in stratum h . Alternatively, $\hat{\gamma}$ may be replaced by the mean of the jackknife estimates, $\hat{\gamma}^J = \sum_{h=1}^H \sum_{s=1}^{k_h} \hat{\gamma}_{(sh)} / \sum_{h=1}^H k_h$. The foregoing procedure also can be used without grouping the PSUs, treating each PSU as a group of size 1.

In general, resampling methods are applied to PSUs without attention to the form of subsampling within the PSUs. This convenient feature is justified by the fact that when the first-stage sampling fraction remains low (<10% for practical purposes), the standard error may be accurately estimated from the variation between PSU totals. The contribution from second and later stage variances is reflected in the sampling error estimated from the PSUs (Lee, Forthofer, and Lorimor 1986). In addition, jackknife variance estimation correctly estimates the component of variance due to sample weighting.

We are unaware of any commercial LC software appropriate for analyzing complex sample survey data. Weighted estimates of LC parameters are provided by the computer packages LEM (Vermunt 1997) and Latent Gold (Vermunt and Magidson 2000); however, these programs do not provide correct estimates for the standard errors for surveys with stratification and clustering. For the current study, GAUSS (version 3.5) (Aptech Systems, Inc. 1997) programming code was written to perform the LCA and the jackknife and verified for traditional LCMs before being applied to complex survey data.

4. ANALYSIS OF DATA FROM THE CONTINUING SURVEY OF FOOD INTAKES BY INDIVIDUALS

We fit a two-class LCM to the CSFII data taking sample weights into account. Three-class models were not assessed for these data, because the unrestricted three-class model is not identified for four variables (Lindsay, Clogg, and Grego 1991). As shown in Table 2, $\hat{\theta}$, the proportion in the first latent class (LC1) is estimated to comprise 18% of the population. LC1 can be interpreted as consisting of “nonregular,” or infrequent, vegetable eaters, that is, those who do not consume vegetables on a regular (daily) basis. The second latent class (LC2), comprising 82% of the population, can be interpreted as consisting of those individuals who consume at least one vegetable as more or less a regular (daily) practice. In LC1, estimates of the item-conditional probabilities for vegetable consumption on a given recall day, $\hat{\alpha}_{1j}$, were variable, ranging from .28 to .46 for vegetable consumption on the j th day, whereas in LC2 these probabilities, $\hat{\alpha}_{2j}$, were similar and consistently higher, ranging from .73 to .78 (see Table 2). Note that we drop the middle subscript (r) for the item conditional probabilities because the responses have only two levels. The jackknife standard error of the LC proportion, .13, was relatively large, as were jackknife standard errors for the item-conditional probabilities in LC1.

In general, the larger the LC, the more observations it represents and the smaller the variability in the estimates for the item-conditional probabilities for that class. We calculated estimates of standard errors based on SRS using a weighted Fisher information based on the (weighted) pseudolikelihood,

Table 2. Latent Class Analysis of Vegetable Consumption Habits: 1985 Continuing Survey of Food Intakes by Individuals

| Parameter | Weighted data | | | Unweighted data | | |
|---------------|---------------|-----------------------------|--------------------------|-----------------|-----------------------------|--------------------------|
| | Estimate | Mean of jackknife estimates | Jackknife standard error | Estimate | Mean of jackknife estimates | Jackknife standard error |
| θ | .178 | .179 | .128 | .331 | .332 | .137 |
| α_{11} | .456 | .456 | .200 | .604 | .604 | .078 |
| α_{12} | .391 | .390 | .227 | .510 | .510 | .094 |
| α_{13} | .275 | .276 | .113 | .396 | .397 | .082 |
| α_{14} | .392 | .392 | .148 | .464 | .464 | .074 |
| α_{21} | .781 | .781 | .021 | .800 | .801 | .019 |
| α_{22} | .764 | .764 | .030 | .818 | .818 | .034 |
| α_{23} | .766 | .766 | .069 | .810 | .811 | .065 |
| α_{24} | .729 | .730 | .040 | .787 | .787 | .046 |

where the weights were normalized to the sample size. These were about one-half the size of the jackknife standard errors. For the LC proportion, the standard error from the Fisher Information was .07, compared to .13 from the jackknife; this translated into a deff of about 4. Deffs for the conditional probabilities ranged from about 1 to 4 (data not shown).

The Akaike information criterion (AIC) has been used to assess goodness of fit for LCMs (Lin and Dayton 1997), but has not been modified for complex sample survey data. We used a Wald test to test goodness of fit for our two-class model, because this test can be adapted to sample survey data by using a design-based estimate of the variance matrix. The Wald test statistic is the quadratic form $W = d'V^{-1}d$, where d is a 15×1 vector of the differences between the observed and expected cell proportions for 15 of the 16 possible outcome cells and V is the estimated variance matrix for d . The jackknife was used to compute V . We compared $W \times (57 - 15 + 1)/(57 \times 15)$, where 57 is the number of PSUs (114) minus the number of strata (57), to an F distribution with 15 and 43 degrees of freedom. (See Korn and Graubard 1999, pp. 91–93, for a discussion of Wald tests.) For the two-class model, the test statistic was .72 ($p = .75$), indicating that the model fits the data satisfactorily. To assess bias in parameter estimates incurred by ignoring sample weights, we fit an unweighted two-class model to the data (see Table 2). For the unweighted data, the estimated proportion falling in LC1 was .33 as opposed to .18 for weighted data. Differences for the conditional probabilities were smaller. Overall, the variances tended to be greater when the weights were used than when they were ignored.

We used Wald tests for the difference between the weighted and unweighted estimates to assess whether the sample weights were informative. Because weighted analyses tend to increase the variance of estimated parameters, these tests are known to have low power. Testing the 8 item-conditional probabilities, the F value for the Wald test with 8 and 57 degrees of freedom was 1.02 ($p = .49$). Testing only the LC proportion, the F value for the Wald test with 1 and 57 degrees of freedom was 2.01 ($p = .16$). The results of this analysis suggest that the weights may not be informative.

The U. S. Department of Agriculture computed a single estimated overall deff of 1.43 for analyzing the 4 days of records for the 1985 CSFII. It was computed as $1 + \{cv(wts)^2\}$ (Joseph Goldman, personal communication), where $cv(wts)$ is the coefficient of variation of the sampling weights. This deff takes into account the variability associated with the weights, but not the effects of clustering or stratification.

As we had decided to retain the weights, we were interested in obtaining an estimate of the deff due to clustering and stratification alone, apart from that due to the weights. For the sample of 1,028 women, we generated a vector of 1,028 uniform random numbers, each number associated with an observation. Next, we randomly regrouped the response vectors into clusters retaining the original cluster sizes. We then fit the reordered data to a two-class model and used the resulting variances to estimate the deff for each parameter estimate. The deff was estimated as .97 for the LC proportion and ranged from .98 to 1.17 for the item-conditional probabilities

Table 3. Latent Class Analysis of Vegetable Consumption Habits: 1985 Continuing Survey of Food Intakes by Individuals Weighted Data, Clusters Broken by Random Reordering

| Parameter | Estimate | Mean of jackknife estimates | Jackknife variance estimate, without clusters | Jackknife variance estimate, with clusters | Ratio of variances with:without clusters |
|---------------|----------|-----------------------------|---|--|--|
| θ | .178 | .179 | .017 | .016 | .972 |
| α_{11} | .456 | .455 | .041 | .040 | .976 |
| α_{12} | .391 | .390 | .050 | .052 | 1.027 |
| α_{13} | .275 | .276 | .011 | .013 | 1.157 |
| α_{14} | .392 | .391 | .020 | .022 | 1.111 |
| α_{21} | .781 | .781 | .000 | .000 | .980 |
| α_{22} | .764 | .764 | .001 | .001 | 1.095 |
| α_{23} | .766 | .766 | .005 | .005 | 1.012 |
| α_{24} | .729 | .730 | .001 | .002 | 1.122 |

(Table 3). These effects were modest compared to the deffs that incorporate weighting as well as clustering, indicating that most of the increase in variance was due to the sample weights.

5. SIMULATION

We performed a simulation to investigate the validity of the methods used for taking weights and clustering into account for the CSFII data and to assess the accuracy of the jackknife standard errors. This simulation was based on numbers of strata (i.e., 60) and PSUs (i.e., 2 per stratum) similar to those in the CSFII. The size of the PSUs in the simulation was set at 8, the average PSU size in the CSFII. For simplicity, all PSUs were of equal size and the sample size was set at 960, a multiple of 8 and similar to the CSFII sample size. A population with an underlying two-class structure was simulated. We drew the LC proportions for 30 of the strata from a beta distribution $\beta(1, 9)$, with mean .1, (i.e., $\theta_1 = .1$), and drew the proportions for the remaining 30 strata from a beta distribution $\beta(3, 7)$, with mean .3 (i.e., $\theta_2 = .3$), so that the proportion in LC1 in the overall simulated population, .2, was close to .18, as estimated in the two-class solution for the CSFII data. We randomly generated values of the LC proportions from these beta distributions, inducing intracluster correlations within PSUs. We selected the beta distribution because it is a flexible two-parameter distribution (scale and location parameters), has values lying in the [0, 1] interval, and is the natural conjugate prior distribution for the binomial distribution. In theory, the intraclass correlation coefficient for a beta distribution with parameters ν, ω is $(\nu + \omega + 1)^{-1}$ (Brier 1980). We set the item-conditional probabilities at .2 for LC1 and .7 for LC2 to approximate the CSFII values.

A plot of the sample weights from the CSFII suggests that they are approximately lognormal in distribution. We used moments of the empirical distribution of the weights to define a lognormal distribution with a median of .84 and a variance of .616, and generated sample weights for the observations in the simulation from this distribution. The simulation can be viewed as a series of one-stage cluster samples where each cluster consists of b observations and where the LC proportions vary by cluster within each stratum.

Table 4. Estimated True and Jackknife Variances for Simulation

| Parameter | Estimated true variance | Jackknife estimate of variance | Ratio of jackknife:true variance |
|---------------|-------------------------|--------------------------------|----------------------------------|
| θ | .00202 | .00217 | 1.074 |
| α_{11} | .00518 | .00534 | 1.031 |
| α_{12} | .00507 | .0054 | 1.065 |
| α_{13} | .00512 | .00532 | 1.039 |
| α_{14} | .00499 | .00532 | 1.066 |
| Mean | .00509 | .00535 | 1.050 |
| α_{21} | .00076 | .00078 | 1.026 |
| α_{22} | .00077 | .00078 | 1.013 |
| α_{23} | .00075 | .00077 | 1.027 |
| α_{24} | .00077 | .00079 | 1.026 |
| Mean | .00076 | .00078 | 1.023 |

To investigate the effect of clustering on the jackknife variance, we generated clustered data from the aforementioned population, and estimated the variance taking the clustering into account. We then calculated the jackknife variance for a sample from the same population constructed using the reordering method described in the previous section in the discussion of the deff. We randomly regrouped observations into clusters of the same size and, using these clusters as PSUs, estimated jackknife variances.

The code for the simulations was written in the matrix language, GAUSS, version 3.5, and the EM algorithm was used to estimate model parameters. The programming criteria used in the simulation were (1) 1,000 replications, (2) convergence criterion of 10^{-6} , and (3) maximum number of 500 iterations allowed to achieve convergence in the LCA algorithm (non-converging cases were replaced in the simulation).

To assess the validity of the jackknife variances from the simulations, we generated proxy population variances by calculating mean squared errors for the parameter estimates based on 10,000 replications using the same parameter values as in the simulations. The ratio of the jackknife variance estimate to the corresponding proxy variance was taken as a measure of the accuracy of the jackknife estimate. A 95% two-tailed confidence interval (CI) was calculated for the simulation parameter values as

$$CI = \left\{ \hat{\theta} - t_{\frac{\alpha}{2}, df} \sqrt{\widehat{\text{var}}^J}, \hat{\theta} + t_{\frac{\alpha}{2}, df} \sqrt{\widehat{\text{var}}^J} \right\}, \tag{7}$$

where $\hat{\theta}$ can be either the LC proportion or an item-conditional probability, α is the type I error rate, df is the number of (jackknifed) groups minus the number of strata, and $\widehat{\text{var}}^J$ is the jackknife variance estimate.

As expected (Kish and Frankel 1974), for all parameters, the jackknifed variances modestly overestimated the proxy variances (Table 4). Estimates for the item-conditional probabilities were within 7% of the proxy variances for the smaller LC and within 3% for the larger class. The jackknife overestimated the variance of the LC proportion by 7%. As shown in Table 5, coverage was close to the nominal .95 level for all parameters.

Table 5. 95% Confidence Interval Coverage for Simulation, $n = 960$, $\theta_1 \sim \beta(1, 9)$, $\theta_2 \sim \beta(3, 7)$, Lognormal Weights

| | Proportion in lower tail | Proportion in upper tail | Coverage |
|---------------|--------------------------|--------------------------|----------|
| θ | 0 | .060 | .940 |
| α_{11} | .025 | .042 | .933 |
| α_{12} | .023 | .037 | .940 |
| α_{13} | .034 | .042 | .924 |
| α_{14} | .024 | .032 | .944 |
| Mean | .027 | .038 | .935 |
| α_{21} | .014 | .028 | .958 |
| α_{22} | .017 | .026 | .957 |
| α_{23} | .016 | .030 | .954 |
| α_{24} | .028 | .026 | .946 |
| Mean | .019 | .028 | .954 |

6. DISCUSSION

Multiple dietary records of food intake typically have been summarized by means and proportions. LCA is a new method of combining records to group respondents into categories, or classes, that define patterns of food consumption and provide estimates of class size. We fit an unconstrained model because of the possibility that seasonality or other variables might affect vegetable consumption over the course of the survey year. Fitting a two-class model, we found that about 18% of the population of women age 19–50 consumed a diet deficient in vegetables in that they did not make consumption of these foods a regular practice. LCA also provides estimates of the item-conditional probabilities (class-specific dietary propensity scores). There was a suggestion that respondents tended to be more likely to report consuming at least one vegetable on the first survey day than on later recall days. Because vegetable consumption is advocated as part of a good diet, respondents may have been more likely to report eating a vegetable in the face-to-face interview than when queried by telephone. Although the similarity of the item-conditional probabilities for recalls 2–4, especially for LC2, suggested that a model restricting these probabilities to be equal might be appropriate, we rejected this course because it would have been a *post hoc* analysis. The similarity of the item-conditional probabilities over the 4 recall days for LC2 suggested a stable propensity to consume vegetables. This was not true for LC1.

In this study, we used LCA to estimate the proportion of women age 19–50 that consume vegetables on a regular basis, a different objective than estimating the number of servings per day as in some other types of analysis. LCA requires only indicators of consumption and can lead to data reduction in some datasets. Thus LCA can be readily performed on data that otherwise may require a multiple-step, perhaps lengthy, analysis involving transformations and distributional assumptions. For example, Nusser et al. (1996) proposed a complex multistep procedure for estimating the distribution of nutrient intake. Finally, LCA provides a new way to describe “usual” dietary intake and to estimate the number and size of subgroups that display different food consumption patterns. Such analyses may be useful in developing public health intervention programs.

There has been a long-standing debate in the statistical literature on whether to do weighted or unweighted analysis (i.e., design-based or model-based analysis) of survey data (Brewer and Mellor 1973; Smith 1976, 1984; Hansen, Madow, and Tepping 1983; Fienberg 1989; Kalton 1989; Korn and Graubard 1995a, 1995b). It is well known that using sample weights will result in approximately unbiased or consistent estimates for population parameter values, but may increase the variances of these estimates, whereas an unweighted analysis may result in biased or inconsistent estimates, but smaller variances. We have described a weighted analysis that uses weighted pseudolikelihood estimation, and applied this method to the analysis of CSFII data. Consistency of weighted estimates is maintained regardless of whether the posited model is correctly specified. In contrast, when the sample weights are informative for the analysis of interest, unweighted estimates will depend on the particular sample weighting scheme used in that analysis (Pfefferman 1993). Issues to be considered when choosing a weighted versus an unweighted analysis are (1) the purpose of the analysis—analytical versus descriptive; (2) the magnitude of the inefficiency that would result from a weighted analysis if the weighting were unnecessary to correct for bias and whether this inefficiency is small relative to the effect being estimated; (3) the expected bias from an unweighted analysis; and (4) whether sufficient information is known about the sample design and whether variables are available to model the sample design in an unweighted analysis (Korn and Graubard 1999, chap. 4).

For LC modeling, sample weighting can affect the estimation of the item-conditional probabilities, the LC proportions, or both when sampling rates differ across subgroups. Consider an unstratified analysis of a population comprising two subgroups (i.e., a single LC model fitted to the entire population), where both subgroups have the same number of underlying latent classes but are sampled at different rates. If the LC proportions differ between subgroups, then sample-weighted estimates of the LC proportions will differ from unweighted estimates. If the item-conditional probabilities are homogeneous across subgroups, then the weighted and unweighted estimates of the item-conditional probabilities should be approximately the same, whereas the LC proportions could differ. If these probabilities differ between subgroups, then again weighted and unweighted estimates can differ. If the analysis is stratified so that a separate LCM is fitted to each subgroup, then weighting is no longer necessary. However, stratifying among all population subgroups is rarely feasible.

The CSFII data analysis was a descriptive analysis that used LC modeling without covariates. The objective of the analysis was to obtain unbiased estimates of the LC proportions and item-specific probabilities for the target population. Following the recommendations of Korn and Graubard (1999, pp. 180–182) we used a weighted analysis for this descriptive study. For an analytical study, the trade-off between variance and bias must be carefully considered. An analyst choosing to use unweighted analysis because of large inefficiency due to the weighting should adjust for the sample weighting by including in the analytic model those sample design variables used in determining the sample weighting (Korn and Graubard

1999). Regardless of the type of analysis done, model adequacy should be assessed using diagnostic methods, as we have tried to do here.

Analyses of the CSFII data and the simulations done with and without sample weights demonstrated both the possibility of incurring unacceptable bias by ignoring the weights and the potential increase in variance arising from including them unnecessarily. The CSFII design is described as self-weighting, although weights were used to adjust for eligibility within the household and for nonresponse. The self-weighting aspect of the design might lend support to the notion that the weights could be ignored, although this is not obvious for the present analysis. The Wald test comparing the weighted estimates to the unweighted estimates showed a larger, but not significant, effect on the LC proportions than on the item-conditional probabilities. This test has low power, however.

The jackknife is an easily applied method for obtaining empirical variance estimates for an LCM applied to complex sample survey data. Our simulation suggested that the jackknife standard errors slightly overestimate the actual standard errors. Despite this overestimation, these estimates seem sufficient for most practical applications. However, it may be worthwhile to investigate other resampling methods, such as the bootstrap or modifications to the jackknife. Another proposed approach uses linearization variances based on Taylor series approximations of the estimating equations from the sample weighted pseudolikelihood (Wedel et al. 1998). This approach is less flexible in that it requires developing new software (e.g., for the calculation of second derivatives for each term in the model for each model considered).

Another approach to analyzing data from cluster samples is using hierarchical modeling with random effects to model the correlation at each stage of cluster sampling. The use of random-effects models applied to survey data is an area of current research with no well-established methods, even in the case of linear models (Korn and Graubard 1998; Pfefferman, Skinner, Holmes, Goldstein, and Rasbash 1998). This approach is difficult to implement because it requires knowledge of all levels of clustering, which is often unavailable on public use files because of confidentiality concerns. The approach that we have taken, (weighted) pseudolikelihood with design-based jackknife variance estimation, is commonly used to analyze survey data with complex sample designs (Skinner, Holt, and Smith 1989; Korn and Graubard 1999, p. 101).

We do not recommend inflating the variance by an overall survey deff, as done by Haertel (1984a, 1984b, 1989). First, we found that the jackknife standard errors, which take the sample design fully into account, were about twice as large as standard errors based on Fisher information for a sample-weighted likelihood; this difference translates into deffs of approximately 4. These very large deffs were due primarily to the effects of sample weights, with only modest effects due to clustering and stratification. These deffs varied by parameter and were larger than the overall deff of 1.43 estimated by the U. S. Department of Agriculture.

National surveys such as the CSFII, the National Health and Nutrition Examination Survey, and the National Health Interview Surveys are major sources of information on dietary

practices in the general population and in demographic subgroups. In the past, LCA has not been applied to these data. With the development of methods to accommodate weighted and clustered data, LCA can be used to describe food consumption patterns in the whole population, as well as in subsets of interest.

The 1994–1996 CSFII collected only two 24-hour recalls. LCA can be applied to surveys such as this by introducing two or more latent variables, such as separate latent indicators of fruit and of vegetable intake, and fitting models with two or more classes. These may be independent or correlated, as discussed by Hagennars (1990). Alternatively, multiple group analyses can be performed, where the groups relate, to say, sex, to race, or to some other classification variable (Dayton 1999).

An area of future research is the development of goodness-of-fit test statistics for LC models for survey data. Although sample weights might be readily incorporated into statistics based on the log-likelihood, the distribution of test statistics such as the AIC must be modified to take into account clustering or stratification.

[Received January 2000. Revised April 2002.]

REFERENCES

- Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.
- Aptech Systems, Inc. (1997). *The GAUSS System* (Version 3.5). Maple Valley, WA.
- Block, G., Patterson, B. H., and Subar, A. (1992). "Fruit, Vegetables, and Cancer Prevention: A Review of the Epidemiological Evidence." *Nutrition and Cancer*, 18, 1029.
- Bolesta, M. S. (1998). "Comparison of Standard Errors Within a Latent Class Framework Using Resampling and Newton Techniques." doctoral dissertation, University of Maryland, College Park.
- Brewer, K. R. W., and Mellor, R. W. (1973). "The Effects of Sample Structure on Analytic Surveys." *The Australian Journal of Statistics*, 15, 145–152.
- Brier, S. S. (1980). "Analysis of Contingency Tables Under Cluster Sampling." *Biometrika*, 67, 591–596.
- Clogg, C. C., and Eliason, S. R. (1987). "Some Common Problems in Log-Linear Analysis." *Sociological Methods & Research*, 16, 8–44.
- Clogg, C. C., and Goodman, L. A. (1984). "Latent Structure Analysis of a Set of Multidimensional Tables." *Journal of the American Statistical Association*, 79, 762–771.
- (1985). "Simultaneous Latent Structure Analysis in Several Groups," in *Sociological Methodology 1985*, ed. N. B. Tuma, San Francisco: Jossey-Bass.
- Dayton, C. M. (1999). *Latent Class Scaling Analysis*. Thousand Oaks, CA: Sage.
- Dayton, C. M., and Macready, G. B. (1976). "A Probabilistic Model for Validation of Behavioral Hierarchies." *Psychometrika*, 41, 189–204.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum Likelihood Estimation From Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society*, 39, 1–22.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Methods*. Philadelphia: Society for Industrial and Applied Mathematics.
- Fienberg, S. E. (1989). "Modeling Considerations: Discussion From a Modeling Perspective," in *Panel Surveys*, eds. D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh. New York: Wiley, pp. 566–574.
- Frankel, M. R. (1971). *Inference From Survey Samples*. Ann Arbor, MI: Institute for Social Research, University of Michigan.
- Goodman, L. A. (1974). "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika*, 61, 215–231.
- Guenther, P. M. (1997). "Development of an Approach for Estimating Usual Nutrient Intake Distributions at the Population Level." *Journal of Nutrition*, 127, 1106–1112.
- Haberman, S. J. (1979). *Analysis of Quantitative Data*, Vol. 2. New York: Academic Press.
- Haertel, E. (1984a). "An Application of Latent Class Models to Assessment Data." *Applied Psychological Measurement*, 8, 333–346.
- (1984b). "Detection of a Skill Dichotomy Using Standardized Achievement Test Items." *Journal of Educational Measurement*, 21, 59–72.
- (1989). "Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items." *Journal of Educational Measurement*, 26, 301–321.
- Hagennars, J. A. (1990). *Categorical Longitudinal Data*. Newbury Park, CA: Sage.
- Haines, P. S., Hungerford, D. W., Popkin, B. M., and Guilkey, D. K. (1992). "Eating Patterns and Energy and Nutrient Intakes of U. S. Women." *Journal of the American Dietetic Association*, 92, 698–704, 707.
- Hansen, M. H., Madow, W. N., and Tepping, B. J. (1983). "An Evaluation of Model Dependent and Probability-Sampling Inferences in Sample Surveys." *Journal of the American Statistical Society*, 78, 776–793.
- Hartman, A. M., Brown, C. C., Palmgren, P. P., Verkasalo, M., Myer, D., and Virtamo, J. (1990). "Variability in Nutrient and Food Intakes Among Older Middle-Aged Men: Implications for Design of Epidemiologic and Validation Studies Using Food Recording." *American Journal of Epidemiology*, 132, 999–1012.
- Heinen, T. (1996). *Latent Class and Discrete Trait Models*. Advanced Quantitative Techniques in the Social Sciences Series, Vol. 6. Thousand Oaks, CA: Sage.
- Hidiroglou, M. A., and Rao, J. N. K. (1996). "Chi-Squared Tests With Categorical Data From Complex Surveys: Part I." *Journal of Official Statistics*, 3, 117–132.
- Human Nutrition Information Service (1983). "Food Intakes: Individuals in 48 States, Year 1977–78, Nationwide Food Consumption Survey 1977–78." Report I-1, U. S. Department of Agriculture.
- Kalton, G. (1989). "Modeling Considerations: Discussion From a Survey Sampling Perspective," in *Panel Surveys*, eds. D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh. New York: Wiley, pp. 575–585.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Kish, L., and Frankel, M. P. (1974). "Inference From Complex Samples." *Journal of the Royal Statistical Society, Ser. B*, 36, 1–37.
- Korn, E. L., and Graubard, B. I. (1995a). "Analysis of Large Health Surveys: Accounting for the Sampling Design." *Journal of the Royal Statistical Society, Ser. A*, 158, 263–295.
- (1995b). "Examples of Differing Weighted and Unweighted Estimators From a Sample Survey." *The American Statistician*, 49, 291–295.
- (1998). Discussion of "Weighting for Unequal Selection Probabilities in Multilevel Models," by D. Pfeffermann et al., *Journal of the Royal Statistical Society, Ser. B*, 60, 23–40.
- (1999). *Analysis of Health Surveys*. New York: Wiley.
- Krebs-Smith, S. M., Cook, A., Subar, A. F., Cleveland, L., and Friday, J. (1995). "U. S. Adults' Fruit and Vegetable Intakes, 1989 to 1991: A Revised Baseline for the Healthy People 2000 Objective." *American Journal of Public Health*, 85, 1623–1629.
- Krebs-Smith, S. M., and Kantor, L. S. (2001). "Choose a Variety of Fruits and Vegetables Daily: Understanding the Complexities." *Journal of Nutrition*, 131, 487–501S.
- Lazarsfeld, P. F., and Henry, N. W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.
- Lee, E. S., Forthofer, R. N., and Lorimor, R. J. (1986). "Analysis of Complex Sample Survey Data: Problems and Strategies." *Sociological Methods & Research*, 15, 69–100.
- Lin, T. H., and Dayton, C. M. (1997). Model Selection Information Criteria for Non-Nested Latent Class Models." *Journal of Educational and Behavioral Statistics*, 22, 249–264.
- Lindsay, B., Clogg, C. C., and Grego, J. (1991). "Semiparametric Estimation in the Rasch Model and Related Models, Including a Simple Latent Class Model for Item Analysis." *Journal of the American Statistical Association*, 86, 96–107.
- Miller, R. G. (1968). "Jackknifing Variances." *Annals of Mathematical Statistics*, 39, 567–582.
- (1974). "The Jackknife—A Review." *Biometrika*, 61, 1–14.
- Mosteller, F., and Tukey, J. W. (1968). "Data Analysis. Including Statistics," in *The Handbook of Social Psychology*, Vol. 2, eds. G. Lindsay and E. Aronson. Reading, MA: Addison-Wesley, pp. 80–203.
- Nusser, S. M., Carriquiry, A. L., Dodd, K. W., and Fuller, W. A. (1996). "A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions." *Journal of the American Statistical Association*, 91, 1440–1449.
- Patterson, B. H. (1998). "Latent Class Analysis of Sample Survey Data." doctoral dissertation, University of Maryland, College Park.
- Patterson, B. H., Block, G., Rosenberger, W. F., Pee, D., and Kahle, L. L. (1990). "Fruit and Vegetables in the American Diet: Data From the NHANES II Survey." *American Journal of Public Health*, 80, 1443–1449.

- Patterson, B. H., Harlan, L. C., Block, G., and Kahle, L. (1995), "Food Choices of Whites, Blacks and Hispanics: Data From the 1987 National Health Interview Survey," *Nutrition and Cancer*, 23, 105-119.
- Pfeffermann, D. (1993), "The Role of Sampling Weights When Modeling Survey Data," *International Statistical Review*, 61, 317-337.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., and Rasbash, J. (1998), "Weighting for Unequal Selection Probabilities in Multilevel Models," *Journal of the Royal Statistical Society, Ser. B*, 60, 23-40.
- Popkin, B. M., Siega-Riz, A. M., and Haines, P. S. (1996), "A Comparison of Dietary Trends Among Racial and Socioeconomic Groups in the United States," *New England Journal of Medicine*, 335, 716-720.
- Quenouille, M. H. (1949), "Approximate Tests of Correlation in Time-Series," *Journal of the Royal Statistical Society, Ser. B*, 11, 68-84.
- Roberts, G., Rao, J. N. K., and Kumar, S. (1987), "Logistic Regression Analysis of Sample Survey Data," *Biometrika*, 74, 1-12.
- Skinner, C. J., Holt, D., and Smith, T. M. F. (eds.) (1989), *Analysis of Complex Surveys*, New York: Wiley.
- Smith, A. F. (1991), "Cognitive Processes in Long-Term Dietary Recall," National Center for Health Statistics. Vital Health Statistics.
- Smith, P. J., Graubard, B. I., and Midthune, D. N. (1997), "Mixed Poisson Model for Clustered Count Data With Covariates and Application to Dietary Intake Data," unpublished manuscript.
- Smith, T. M. F. (1976), "The Foundations of Survey Sampling: A Review" (with discussion), *Journal of the Royal Statistical Society, Ser. A*, 139, 183-204.
- (1984), "Present Position and Potential Developments: Some Personal Views—Sample Surveys" (with discussion), *Journal of the Royal Statistical Society, Ser. A*, 147, 208-221.
- Subar, A. F., Frey, C. M., Harlan, L. C., and Kahle, L. (1993), "Differences in Reported Food Frequency by Season of Questionnaire Administration: The 1987 National Health Interview Survey," *Epidemiology*, 5, 226-233.
- Subar, A. F., Heimendinger, J., Patterson, B. H., Krebs-Smith, S. M., Pivonka, E., and Kessler, R. (1994), "Fruit and Vegetable Intake in the United States: The Baseline Survey of the Five A Day for Better Health Program," *American Journal of Health Promotion*, 9, 352-360.
- Thompson, F. E., and Byers, T. (1994), "Dietary Assessment Manual," *The Journal of Nutrition*, 24, 2245S-2317S.
- U. S. Department of Agriculture, Agricultural Research Services (1998), *Food and Nutrient Intakes by Individuals in the United States, by Age and Sex, 1994-1996*, Nationwide Food Surveys Report 96-2, Beltsville, MD: U. S. Department of Agriculture.
- U. S. Department of Agriculture, Human Nutrition Research Service (1987), *CSFII: Nationwide Food Consumption Survey. Continuing Survey of Food Intakes by Individuals, Women 19-50 and Their Children 1-5 Years, 4 Days, 1985*, CSFII Report 85-4, Washington DC: U. S. Government Printing Office.
- U. S. Department of Health and Human Services (2000), *Tracking Healthy People 2010*, Washington DC: U. S. Government Printing Office.
- Vermunt, J. K. (1997), "The LEM User Manual," WORC paper, Tilburg University, The Netherlands.
- Vermunt, J. K., and Magidson, J. (2000), *Latent Gold: User's Guide*, Belmont, MA: Statistical Innovations, Inc.
- Wedel, M., ter Hofstede, F., and Steenkamp, J. E. M. (1998), "Mixture Model Analysis of Complex Surveys," *Journal of Classification*, 15, 225-244.
- Wolter, K. M. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.
- Young, L. R., and Nestle, M. (1995), "Portion Sizes in Dietary Assessment: Issues and Policy Implications," *Nutrition Reviews*, 53, 149-158.

Comment

Alicia L. CARRIQUIRY and Sarah M. NUSSE

It is well known that collecting and analyzing dietary intake data can be challenging (e.g., Beaton et al. 1979; Basiotis, Welsh, Cronin, Kelsay, and Mertz 1987; Dwyer and Coleman 1997). Yet despite the difficulties inherent in accurately measuring food intakes and in drawing useful inferences from those measurements, the U. S. government relies on complex dietary intake surveys to guide nutrition and health policy, monitor the performance of food assistance programs, and design interventions such as national food fortification programs. In this light, the work of Patterson, Dayton, and Graubard is welcome in that it seeks to capitalize on the rich data available for dietary assessment.

In this discussion we focus on the subject matter interpretation and the statistical aspects of the variable used to indicate dietary intake as well as the latent class model used to make inferences using this variable. In the next section we discuss the importance of informative dietary intake measures with respect to policy development. In Section 2 we focus on the model itself. Finally, in Section 3 we provide some conclusions and thoughts.

1. VARIABLES OF INTEREST TO POLICY MAKERS

In nationwide surveys such as the Continuing Survey of Food Intakes by Individuals (CSFII), respondents are asked to report on the amounts of food consumed during the previous 24 hours. The amounts of the various foods consumed are expressed in such units as glasses, cups, grams, slices, tablespoons, and so forth. Even though the interviewer arrives at a respondent's home armed with two- and three-dimensional models that are meant to help the respondent to accurately quantify the amount of each food consumed, it is still well known that correctly gauging portion sizes can be difficult (Hartman et al. 1994; Haraldsdottir, Tjonneland, and Overvad 1994; Dwyer and Coleman 1997). When interviews are conducted over the phone, measurements are likely to be even more inaccurate. In this sense, the authors correctly argue that the measurement error in dietary intake data can be significant. They believe that the presence of this measurement error, compounded by the fact that respondents tend to underreport

Alicia L. Carriquiry is Professor, and Sarah M. Nusser is Associate Professor, Department of Statistics, Iowa State University, Ames, IA, 50011 (E-mail: alicia@iastate.edu and nusser@iastate.edu).